A COMPARATIVE ANALYSIS OF APPROACHES TO DESIGN AND CAPITALIZE DATA INDICATORS

Albane Gril^{1,2}, Madeth May¹, Valérie Renault² and Sébastien George¹ Le Mans University, ¹LIUM-EA4023, ²CREN-EA2661, France

ABSTRACT

In Technology Enhanced Learning field, learning analytics cover multiple research challenges, among which tracking data analysis and data indicator design and visualization. Part of our research effort is dedicated to changing their design process, in order to capitalize them. This would allow us to meet a need in cost savings of design workflow and to encourage the adoption by users throughout their implications in a simplified design process. A study of the state of the art has been made to explore various solutions which could be exploited in the capitalization of indicators on both levels, design and implementation. It sets a direction toward a user-centered approach, aiming at a better control of users over their observation needs and the use of their data.

KEYWORDS

Learning Analytics, Capitalization, Indicators, Traces, User-Centered Design, Comparative Analysis

1. INTRODUCTION

In a context where the use of digital environments has become unavoidable, educational practitioners use TEL systems (Technology Enhanced Learning) on a daily basis, generating at the same time a massive quantity of traces. Numerous works exploit these traces with data analysis and visualization mechanisms, transforming raw traces into visual data indicators that are much more comprehensible to users. Our research effort falls into this research area and has been carried out along with a multi-party university project since 2018. The main objective of the latter is to promote students' success via a learning platform to support them in developing their skills in French writing. Our research aims to study how to provide the participants in the project with a technological solution built upon trace analysis and data indicator visualization.

An indicator, according to Dimitrakopoulou (2004), is "a variable in the mathematical sense to which a series of characteristics is assigned". In a pedagogical context, it is an effective way for both teachers and students to take full possession of this digital dimension of learning, as shown by Alowayr and Badii (2014). The design process can be broken down into three steps: defining the observation needs, computing data indicators and data exploitation. However, each of these steps is highly dependent on the context. For example, the environment and the user target the data indicators are designed for, have an important influence on the definition of the indicator (Papamitsiou *et al.* 2012). The intervention of an analyst in the process is therefore necessary. It makes their design resource-intensive. Given this complexity, it is interesting to limit the repetition of the whole process to each new need or context in order to make the design more cost-effective. There is also a limit to their adoption. A feeling of inadequacy of the proposed indicators in relation to the requirements of the end users is sometimes noticed. The difficulty of identifying the observation needs of the various users and the difficulty of extracting the information can be a deterrent (Littlejohn *et al.* 2009).

An approach based on the capitalization of indicators aims to respond to the need in cutting down the design costs, thus enabling a better use of both technical and human resources. We seek to develop a process for the design of indicators that allows their capitalization by the users of the TEL system. Indeed, by allowing a simplification of the design process, the direct collaboration of users is facilitated.

To meet this goal, the issue has been approached using two complementary methods. To initiate a co-design process with users, a protocol for collecting needs is being validated. This is meant to better grasp the involvement users can, and want to, have in the workflow. This will be detailed in another article. In parallel, a state of the art of indicator design approaches has been conducted. The study helps us consolidate the

definition of capitalization, before isolating specific challenges to focus on. By using low level criteria to identify indicator capitalization, we can get closer to the user's pragmatic needs. A higher level of interpretation of those requirements will, then, allow us to confront our primary conception of capitalization in a process of consolidating the definition of the notion.

This paper is structured as follows. First, we point out the need for data indicator capitalization and make an attempt to provide a definition of the latter. Then we specify how we established a list of criteria aiming at comparing different approaches of indicator design. Section 2 is dedicated to the description of the approaches chosen for this state of the art. The comparative study is made later in the same section.

2. CAPITALIZATION STUDY

2.1 Defining Data Indicator Capitalization

The concept of capitalization is usually tied to monetary valuation. Here, the capitalization is considered as a valuation of the data itself and the value its information provides. When it comes to capitalized data indicators, their value can be defined by several factors including how relevant an indicator is, or how easy to use and reuse the existing indicators are, according to users and contexts.

Capitalizing indicators would be about keeping and multiplying their value in time, contexts and users, while maintaining the reusability. Our approach considers the latter as objects with challenges to overcome, like production cost or user validation. Indeed, they have a complex design process, highly dependent on context. Even with the many intents to simplify design workflow for analysts, or observation needs' expression for users (Laforcade et al. 2009), the adoption of indicator in learning practice for both teacher and student remains challenging. On top of that, while the usefulness of dashboards has been demonstrated in Verbert et al. (2013) to back up educational practitioners in any type of pedagogical scenario, many barriers remain to fully assimilate indicators as a common pedagogical resource (Surry et al. 2005). The inadequacy of user's implication and expertise with the current solutions have only recently been considered as a design workflow issue. The commitment to create an adequate solution has led to the rise of co-design processes with users (Dollinger *et al.* 2019). Another reflection often left aside resolves around data privacy and ethic of e-learning. Although its impact on users' perception and even performance is not negligible (May et al. 2017). These challenges are those addressed through indicator capitalization.

This approach of capitalization is therefore the combination of several concepts, as illustrated in Figure 1. The first concept is "appropriability", often used in human sciences (Ollagnier-Beldame *et al.* 2010) and in the field of HCI (Human-Computer Interaction). For instance, when adopting the indicators, users can apprehend and customize them without the intervention of experts. Then, comes the idea of "reusability", that is, indicators must be designed to work with similar context. Next, "adaptability" is needed, meaning modifications can be made to other contexts or meet different needs. Finally, "shareability" must be ensured between users, thus, allowing users to access and share capitalized indicators within the system.

The association of all those concepts under the idea of indicator capitalization aims to tackle the identified challenges and encourage a broader use of capitalized data indicators in e-learning practices. In order to verify the achievement of those four concepts, each was broken down into concrete technical properties as validation criteria, as presented hereafter.

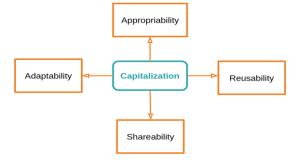


Figure 1. Concepts forming indicator capitalization

2.2 Criteria to Compare Indicator Design Approaches

The criteria of this comparative analysis are intended to specify how an existing approach can achieve indicator capitalization. Several iterations have been made to better define each criterion.

In order to come up with a solution suitable to the platform's users, communication with the project's stakeholders has been maintained right from the start of the design process. Working groups, held primarily with teachers, allowed us to identify some of their needs and expectations. This observation work did not include any presentation of the indicator capitalization process. The subject was tackled through the prism of pedagogical information, their use and challenges, creating a real involvement of the teachers, even without specific knowledge in learning analytics. This will allow us to create a relevant protocol for further experimentations. The preliminary list of needs that emerged was used as a basis to compose a series of technical properties, necessary for the verification of capitalization's concepts (Table 1). The existing criteria are meant to be fed from the experience of different users' context and profiles. The multiple iteration of user need gathering does not affect the existing criteria, but rather help create a hierarchy among the identified needs.

The properties formulated are low level to get closer to functionalities and make the validation process easier. Indeed, an abstract concept is too open to interpretation to be used in a comparative approach. But to formulate technical properties generic enough to apply to different kinds of approaches is no less tedious. Here after a detailed list of all currently selected technical properties classified between the four concepts:

2.2.1 Appropriability

Users need to regain control over the information available to them, which means that anyone can use the system, regardless of specific technical skills.

- **Granularity of entities**: The idea that each complex entity can be broken down into smaller and less complex entities. The more levels of decomposition, the better. This gives the opportunity to consider a nesting architecture to simplify the notions being handled.
- **Visualization choices**: Is an important part of data exploitation, a formalism should be able to specify how an indicator will be visualized, and an implementation should give the possibility to change it. A tool should have customizable visualization options.
- Categorization of entities: By classifying all entities, users will be able to access them more easily. It is even more efficient on multiple levels or with a tag system.
- Search in existing indicators: Users can access and select existing indicators for their own use.
- **Filtering of indicators**: A variety of indicators can be specified and narrowed down, allowing a better choice for the most relevant indicators for each user's need.

2.2.2 Reusability

Any entity created has to be valued within the system, without losing its value through time.

- **Description of entities**: Any information is useful. It can be considered as meta-data of any sort. It is important for meta-data not to be mandatory, but it's availability, updatable by any user, is a plus. This co-construction raises the question of community regulations to maintain a high level of reliability.
- **Knowledge evolution**: Evolution of the existing entities has to be eased. An existing indicator can be modified to create multiple new ones, without losing any previously created entities.
- Versioning system: The idea of keeping track of the indicator's construction through the different steps, as well as its evolution as it passes from user to user. The functionality will allow "Knowledge evolution" without loss of value through time.
- Interlinking of indicators: The possibility to create more complex indicators not only from traces but also from other indicators, all entities should be linkable and those links modifiable. It is an easier way to promote "Knowledge evolution" by using some kind of "Granularity of entities".
- **Storage of entities**: It is necessary to store the existing and created entities to ensure their future reuse. This is not necessarily handled by the tool itself, but needs to be considered when designing a system.

2.2.3 Shareability

To promote community and co-construction, communication between the user of the system can be achieved through different means.

- User access management: To create a system open to many types of user, being able to handle role and permission, on access or edition is necessary. Also, from an ethical perspective, an open system has to consider anonymization of data on distinct levels.
- Circulation of indicators to users: In an effort to co-build knowledge, spreading of the created and existing entities will allow different levels of user commitment and real cost saving. Ideally, any entities or set of entities can be shared to any related group of users.
- Rating of indicators: To promote reuse of existing entities and help newcomers to navigate the potentially great quantity of entities. It can be done in many ways, including assessing different aspects of the indicators, such as their usefulness and their relevance.
- **Comments**: Can be made on any indicator. Users participate through the commenting system in describing indicators and their properties.

2.2.4 Adaptability

The system has to provide a strong support to the expansion of indicator use.

- List of existing entities: It is required for many previous properties to have an available list of all existing entities. An entity includes indicators but also traces, operators, analysis processes or even users.
- Import of external traces: In a new context, traces and data in general, won't necessarily be the same, adding new information has to be facilitated by design.
- Context of the indicator: If the context changes, the one in which an indicator has been created has to be specifically detailed in its description. That context is among other things about the pedagogical situation and the environment' setting of the TEL system.
- **Abstraction of indicators**: Regardless of the context of use, each indicator can be created independently.
- **Modular data computing**: The calculation process needs to be easily adaptable to a new context, if possible automatically by the system, requiring a rigorous formalism set in "Abstraction of indicators".

These technical properties are classified between the four concepts defining capitalization. However, most of this classification can be questioned. For instance, the possibility to search and filter indicators can both enable appropriation, but are related to properties classified in other concepts. In order to search among existing entities, it is necessary to have an available list of the existing ones. This related property "List of existing entities" is placed in the "Adaptability" category. In the meantime, an implementation of a formalism allowing access to existing entities does not necessarily have a search functionality either. Likewise, if the filtering property is strayed forward in a tool, for a formalism it means that indicators have specifications they can be filtered on. Therefore, "Categorized entities" or "Description of entities" are considered related properties without being interdependent.

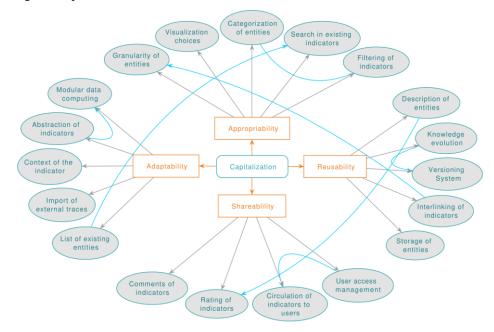


Figure 2. Related criteria for capitalization

Considering the relation of some criteria (Figure 2), their classification was adjusted within multiple iterations of the literature review. The lack of information and inadequate wording was also taken into account to make criteria relevant to the contexts encountered.

With this list of capitalization properties, we seek to experiment how design workflow responds to the identified requirements. Comparing existing solutions through this new perspective and the present grid of criteria will allow us to consolidate the definition of data indicator capitalization within the current context.

3. LITERATURE REVIEW

3.1 Selection of Indicator Design Approaches

Approaches and tools for designing indicators described in various research works have been selected. The starting point was the state of the art carried out in the Thematic Groups of the Digital Education Directorate by Cherigny *et al.* (2020), completed by additional research. Once this initial list of tools had been compiled, those closest to the criteria described above were chosen, distinguishing between formalisms of various kinds and tangible implementations.

Our project's platform is meant for students' online training before a certification, depending on the level reached. It is also meant for teachers to follow the class's progress and help them through available resources. All the approaches selected are relevant in the specific context of our project, removing many tools designed exclusively for interaction analyses, same for laboratory classes and applied work. Likewise, some work specialized on innovative hardware interfaces (tablets, surface tables, VR glasses, ...) were discarded as irrelevant to our work.

Each of these approaches were selected for their relevance in a capitalization of indicator design, but also for the diversity of their nature and solution provided.

Rule-based systems (Schauer 2002) reproduce the cognitive mechanisms of an expert in a particular domain. They are based on facts and modifiable rules, which offers flexibility in data computing and the opportunity for knowledge evolution. It has been implemented in GINDIC (Generator of INDICators). Designers can use this tool to define indicators, the calculations performed on the traces and their visualization (Gendron et al. 2012).

The UTL 2 language (Choquet and Iksal 2007) is an XML-based meta-language describing the traces of a scenario. It uses the description of observation analysis methods to facilitate their capitalization. It can be used for indicator modeling with a strong attachment to pedagogical situations.

The trace-based systems, TBS, (Settouti *et al.* 2009) consider traces as a temporal sequence of observed events along with a capitalizable model. Traces in heterogeneous formats can thus be manipulated to create different indicators abstracted from environment' setting of tools. TBS-IM (Djouad and Mille 2010) is an TBS implemented for the indicator calculation in the Moodle learning platform. Another TBS is DDART (Michel *et al.* 2017), a dynamic dashboard based on activity traces and reports in the form of a Moodle plug-in. Learners can create customizable indicators related to their activities.

The Reusable Indicator Template of David *et al.* (2005) is composed of a computational function of traces and metadata explaining the reuse conditions, including the function's domain of definition, the learning situation and a description. The EM-AGIIR tool (David *et al.* 2005) proposes an implementation of the RIT and an open multi-agent architecture to apply the indicators with new traces.

The ontological framework of Lebis (2018) is used to capitalize the analysis processes of learning traces in a so-called narrative way. It is not meant for indicator capitalization, but its originality motivated us to compare it with the rest. It is part of a larger set of tools built for analysis process capitalization.

The Academic Analytics Tool (Ross *et al.* 2017), was developed for the Moodle Analytics project. AAT is a software allowing users to access and analyse student behaviour data in learning systems. Course Insights is a Learning Analytics Dashboard that provides filterable and comparative visualizations of aggregated students' daily activity and learning events to teachers. (Shabaninejad *et al.* 2020). The flexibility of those tools, which are not based on a studied formalism, are interesting from a user perspective.

We looked into indicators design approaches from the Learning Analytics domain, but it seems relevant to expand this search into other domains where context is equally important. Overall, few tools offer a solution for capitalization by the system as well as the users themselves.

3.2 Literature Review Grid

This state of the art is synthesized in Table 1. Our proposal is based on four levels of validation of a criterion: [V] validated with respect to our work, [PV] partially validated, [NV] not validated and [NR] not relevant, in case of inadequacy with a criterion. The number of "not relevant" criteria is due to the effort to compare objects of different nature with a common grid. The formulation of the criteria as technical properties is interpreted in a distinct way depending on whether the analysed object is a theoretical formalism or a tangible implementation. For instance, "Visualization options" would be the information stored for a language, the display handled for an architecture and real visualizations for implemented tools. In some cases, the property does not apply for some level of abstraction, hence the use of "not relevant".

Validation of each criterion was done through a careful analysis of the literature. Nonetheless, the collected information for our analysis can be limited, as each tool we studied is not always meant for data indicator capitalization. The grid's filling considers interconnection of criteria in the implementation level to fill the information gaps. It may occur that specific information contradicts those supposed interconnections, for instance no filtering implemented even where categorization is permitted.

framework (Lebis 2018 Template David et al. Narrative ontological Settouti et al. 2009 Rule -based System **Trace-base system** Reusable Indicator JTL 2 language (Iksal (Schauer 2002) Choquet 2007) DDART (Michel et al GINDIC (Gendron et (Shabaninejad et al. EM-AGIIR (David et Academic Analytics (Djouad and Mille Implementations Tool (Ross et al. Course Insights TBS-IM system al. 2012) al. 2005) 2010) 2017) 2017) Granularity of entities NV NV Visualization choices NV NV NV Categorization of entities NV NV Search in existing NV indicators Filtering of indicators NV N۷ NV Description of entities PV Knowledge evolution Versioning System NV N۷ NV Interlinking of indicators Storage of entities User access management NV Circulation of indicators to NV NV users Rating of indicators NV NV NV NV NV NV Comments of indicators NV List of existing entities Import of external traces Context of indicator Abstraction of indicators Modular data computing

Table 1. Comparison of formalisms and implementations of indicator capitalization

In this table 1, it can be seen that a rule-based system does not offer any specific means to categorize entities, whereas GINDIC does include the dimension and the nature of the created indicators. On this part, the Reusable Indicator Patron provides three types, Cognitive, Social and Affective, which are here considered as categorization. To take another example, "User access management" may seem a stray forward functionality for an implemented tool, therefore, it can be seen that three out of five of them have it implemented. Then for the formalisms, both rule-based system and trace-based system are considered as partially validating the criterion by suggesting a system architecture in which the property could be easily added, but this concern has not been found in the studied documentation and therefore not fully validated. For the rest of the formalisms, their nature is too disconnected from such a functionality to be considered relevant. A similar reflection has been followed for each criterion in the grid, while studying the available documentation.

Regarding the criteria suggested for the capitalization of indicators, some limits in the design solutions can be noticed. It would be interesting to look for an approach which facilitates the sharing of existing indicators without losing the flexibility of their design. Indeed, the criteria for appropriation and sharing of indicators are the least developed. The usability of an approach is not directly included in this grid, but this important factor of user adoption can be partially assessed through the present criteria.

It can also be noted that an implementation does not necessarily preserve all the technical properties of the corresponding formalism. Some other properties do not seem to be related to the implemented indicator design formalism. However, the chosen formalism clearly has an impact on the implemented functionality. Thus, the conception model for an indicator design approach should not be taken lightly.

4. CONCLUSION

The democratization of data indicators in the learning practice face challenges with high design cost and limited user adoption. Capitalization as a solution to those challenges requires a closer look at the user's needs regarding data collection, transformation, computation, exploration and visualization. The multiplicity of users' level of interaction with the TEL system, the pedagogical information and the data indicator make the design of a common approach more complex. Numerous issues as data privacy and ethics must be assimilated in the design process to reach a thorough user-adapted proposition. The recycling process of an object must be considered at the design stage. However, capitalization also requires the active involvement of users. They need to be proactive for successful indicators' assimilation in the learning practices. This proactivity has to be supported and promoted by the process itself.

With this perspective, a co-design process has been initiated with working groups to identify technical criteria for the capitalization of indicators. These criteria will be a means of validating our approach while relying on a state of the art in the field. This last one shows the necessity to accentuate the accessibility of the entities used by the users to ensure the durability of our approach.

The work presented here allows us to draw some initial analyses, but some limits are to be underlined. First, the domain and project context in which the choice of design approaches remains narrowed. This choice was also due to size constraints, but it would be interesting to extend this analysis to new design solutions. The lack of experimentation to validate the grid's fulling has also been pointed out, and an analysis carried out exclusively through the literature would need to be further investigated. Finally, the criteria developed here are based on a user-centered approach that has only been initiated.

Further work with users will be carried out to validate the initial conclusions. The consolidation of our definition of data indicator capitalization will involve users in the construction of this approach through the organization of focus groups. The grid provided here will allow us to rely on a predefined but enhanceable list of criteria to provide an evolutive approach to indicator design.

ACKNOWLEDGEMENT

Our work is carried out within the French project PIA écri+ (n° ANR-17-NCUN- 0015) http://ecriplus.fr/>.

REFERENCES

- Alowayr, A., Badii, A., 2014. Review of Monitoring Tools for E-Learning Platforms. *International Journal of Computer Science and Information Technology*, 6, 79–86.
- Cherigny, F., El Kechai, H., Iksal, S., Lefevre, M., Labarthe, H., Luengo, V., 2020. L'analytique des apprentissages avec le numérique. *Groupes thématiques de la Direction du numérique pour l'Éducation* (DNE -TN2) (Research Report).
- Choquet, C., Iksal, S., 2007. Modeling Tracks for the Model Driven Reengineering of a TEL System. *Journal of Interactive Learning Research*, 18.
- David, J.-P., Adam, J.-M., Diagne, F., Lejeune, A., Luengo, V., Pernin, J.-P., 2005. Design patterns for recording and analysing usage of learning systems: State of art of tracking and analysing usage.
- Dimitrakopoulou, A., 2004. State of the art on Interaction and Collaboration Analysis.
- Djouad, T., Mille, A., Reffay, C., Mohamed, B., 2009. Engineering of activity indicators using modelled traces for Computer Supported Collaborative Learning. *STICEF*, 16.
- Dollinger, M., Liu, D., Arthars, N., Lodge, J., 2019. Working Together in Learning Analytics Towards the Co-Creation of Value. *Journal of Learning Analytics*, 6, 10–26.
- Gendron, E., Pourroy, F., Carron, T., Marty, J.-C., 2012. Towards a structured approach to the definition of indicators for collaborative activities in engineering design. *Journal of Engineering Design*, 23, 195–216.
- Ji, M., Michel, C., Lavoué, E., George, S., 2014. DDART, a Dynamic Dashboard for Collection, Analysis and Visualization of Activity and Reporting Traces, in: Rensing C., de Freitas S., Ley T., Muñoz-Merino P.J. (Eds) *Open Learning and Teaching in Educational Communities*. Presented at the EC-TEL 2014, Springer, Cham.
- Laforcade, P., Zendagui, B., Barré, V., 2009. Specification of Observation Needs in an Instructional Design Context A Model-Driven Engineering Approach. *CSEDU*, pp. 111–118.
- Lebis, A., Lefevre, M., Luengo, V., Guin, N., 2018. Capitalisation of analysis processes: enabling reproducibility, openness and adaptability thanks to narration. *LAK '18: Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, pp. 245–254.
- Littlejohn, A., Margaryan, A., Milligan, C., 2009. Charting Collective Knowledge: Supporting Self-regulated Learning in the Workplace. *Proceedings of 9th IEEE International Conference on Advanced Learning Technologies*, 208–212.
- May, M., Iksal, S., Usener, C.A., 2017. The Side Effect of Learning Analytics: An Empirical Study on e-Learning Technologies and User Privacy. in: Costagliola, G., Uhomoibhi, J., Zvacek, S., McLaren, B.M. (Eds.), *Computers Supported Education, Communications in Computer and Information Science*. Springer International Publishing, Cham, pp. 279–295.
- Michel, C., Lavoué, E., George, S., Ji, M., 2017. Supporting awareness and self-regulation in project-based learning through personalised dashboards. *IJTEL* 9, 204.
- Ollagnier-Beldame, M., 2010. Footprinting e-learners' activity: A first step to help their appropriation of the training system? *International Journal of Knowledge Management*, 6, 65–77.
- Papamitsiou, Z., Economides, A., 2014. Learning Analytics and Educational Data Mining in Practice: A Systematic Literature Review of Empirical Evidence. *Educational Technology & Society*, 17, 49–64.
- Ross, T., Chang, T.-W., Ives, C., Parker, N., Han, A., Graf, S., 2017. The Academic Analytics Tool: Workflow and Use Cases. *Lecture Notes in Educational Technology*, pp. 231–236.
- Schauer, F., 2002. Playing by the Rules: A Philosophical Examination of Rule-Based Decision-Making in Law and in Life. Oxford University Press.
- Settouti, L.S., Prié, Y., Marty, J.-C., Mille, A., 2009. A Trace-Based System for Technology-Enhanced Learning Systems Personalisation, Proceeding of 9th IEEE International Conference on Advanced Learning Technologies. IEEE, Riga, Latvia, pp. 93–97.
- Surry, D., Ensminger, D., Haab, M., 2005. A model for integrating instructional technology into higher education. *British Journal of Educational Technology*, 36, 327–329.
- Verbert, K., Duval, E., Klerkx, J., Govaerts, S., Santos, J., 2013. Learning Analytics Dashboard Applications. *American Behavioral Scientist*, 57.